

Datenorientierte Inhaltserschließung: Neue Dienste der VZG

Jakob Voß & Uma Balakrishnan
Verbundzentrale des GBV (VZG)



#BiblioCon 2023
2023-05-24



Welche Inhaltserschließung?

- Hier nicht: *Abstracts und freie Schlagwörter*
- Sondern: *Kontrollierte Vokabulare, Klassifikationen, Thesauri...*
 - Universalklassifikationen (BK, RVK, DDC...)
 - Normdateien (GND, LCSH...)
 - Fach und Lokalsystematiken (STW, NSK...)

- Maschinenlesbare, offene Daten
- Web-Standards statt Bibliotheks-Standards

Voraussetzungen



- Jedes Vokabular muss in **BARTOC** eingetragen sein
- Einträge des Vokabulars haben **Identifizier**
- Identifizier haben eine Form als **URI** (können wir einrichten)
- Vokabular ist idealerweise **maschinenlesbar** als Open Data verfügbar

Beispiel: *Basisklassifikation (BK)*

- <http://bartoc.org/en/node/18785>
- **83.71** Handel
- <http://uri.gbv.de/terminology/bk/83.71>

Dienst: Infos über Vokabulare



- **Recherche** von Vokabularen & Eintragung in BARTOC
 - Wer ist für jeweiliges Vokabular verantwortlich?
 - Wo wird das Vokabular genutzt?
 - Wie kann auf das Vokabular zugegriffen werden?
 - ...

- Zunehmend relevant im Bereich Forschungsdaten und Datenintegration
- Ausbau geplant im Rahmen eines beantragten DFG-Projekt
<https://doi.org/10.5281/zenodo.7673393>

Dienst: Konvertierung und Hosting von Vokabularen

- **Konvertierung** von Vokabularen in maschinenlesbare Form
z.B. Excel/CSV → JSKOS/RDF
- **Hosting** der Vokabulare und Bereitstellung per APIs
- Bisher nur bei konkretem Bedarf
- VZG: Vokabular-Management per DANTE (bisher vor allem Museumsbereich)
- Besser: Hosting durch Vokabular-Herausgeber oder andere Terminology Services
 - Erfolgreiche Beispiele: UB Konstanz und UB Bielefeld

Schnittstellen

Zugriff auf Erschließungsdaten

Vorhandene Schnittstellen für den K10plus

<https://wiki.k10plus.de/display/K10PLUS/Schnittstellen>

- **SRU**: Suche per CQL-Suchsyntax
- **unAPI**: Abruf einzelner Datensätze

Nachteile

- Rückgabeformate nur PICA und MARC
- Sacherschließungsdaten über verschiedene Felder verteilt
- Unsaubere Daten

[relativ] neu: JSKOS-API

- Zugriff auf Vokabulare und Mappings
- Format **JSKOS** basiert auf JSON und RDF
- Gleiche API für mehrere Datenbanken
 - **BARTOC**: Informationen über Vokabulare
<https://bartoc.org/api/>
 - **DANTE**: Ausgewählte Vokabulare für Digitale Sammlungen
<https://api.dante.gbv.de/>
 - **coli-conc** : Ausgewählte Vokabulare und Mappings
<https://coli-conc.gbv.de/api/>
 - ...
- Einbindung von Vokabularen und Mappings in weitere Anwendungen (DA-3, WinIBW...)

- **A** Allgemeines
- **B** Theologie und Religionswissenschaften
- ✓ **CA - CK** Philosophie
 - CA 1000** Zeitschriften und Reihenwerke
- **CB** Allgemeines
- ✓ **CC** Systematische Philosophie
 - CC 1000** Versuche zur Grundlegung der Philosophie, Gesamtsysteme
 - CC 1100** Untersuchungen zur Philosophie
 - CC 1200** Geschichte und Untersuchungen philosophischer Methoden
- ✓ **CC 2000 - CC 2600** Logik
 - CC 2000** Geschichte der Logik
 - CC 2200** Werke aus dem Gedankenkreis der traditionellen Logik
 - CC 2400** Einführungen, Handbücher, Sammelbände und Bibliografien

Neu: Occurrences-API

- <https://coli-conc.gbv.de/occurrences/>
- **Occurrences**
 - Anzahl der mit einer Klasse erschlossenen Titel
 - Link auf Titellisten im K10plus-OPAC
 - <https://coli-conc.gbv.de/occurrences/api/?member=http:%2F%2Frvk.uni-regensburg.de%2Fnt%2FQR%2520544>
- **Co-Occurrences**
 - Häufigste gemeinsam vergebenen Notationen

CO Co-Occurrences

RVK	QR 544 Nahrungsmittelleinzelhandel	BK	85.40 Marketing	35
RVK	QR 544 Nahrungsmittelleinzelhandel	BK	83.71 Handel	33
RVK	QR 544 Nahrungsmittelleinzelhandel	BK	83.66 Agrarwirtschaft	11

Neu: Mapping-Inference API

- Ermittlung herleitbarer Sacherschließung auf Grundlage von Mappings (**coli-rich**)

Beispiel

- **RVK QR 544 Nahrungsmittelleinzelhandel** ⇒ **BK 83.71 Handel**
Weil **RVK QR 540 - QR 544 Handelsgewerbe** > broader match **BK 83.71 Handel**
<https://coli-conc.gbv.de/api/mappings/infer?from=http:%2F%2Frvk.uni-regensburg.de%2Fnt%2FQR%2520544&fromScheme=http://bartoc.org/en/node/533&toScheme=http://bartoc.org/en/node/18785>

Mapping-Info

Quellvokabular: RVK Regensburger Verbundklassifikation

Quellkonzept: QR 540 - QR 544 Handelsgewerbe

Zielvokabular: BK Basisklassifikation

Zielkonzept: 83.71 Handel

Mappingtyp: spezifischere Bedeutung (<)

Erstellt von: Melanie Janßen

Erstellt: 2023-05-05

Teil von Konkordanz: [Q Wirtschaftswissenschaften](#) von Melanie Janßen (in [Konkordanzregister öffnen](#))

Identifizier: <https://coli-conc.gbv.de/api/mappings/42dd4f21-a6a2-4701-87df-19bba58d2018> (URI)

urn:jskos:mapping:content:c93a1ac5c1369c4f5066526f12b58a0af5c3337f

urn:jskos:mapping:members:358cc00d97cd8ee00068cd31c923758437ad7e3c

Anreicherung im K10plus <https://opac.k10plus.de/DB=2.299/CMD?>

suchen: <ACT=SRCHA&IKT=8659&TRM=https+coli+conc+gbv+de+api+mappings+42dd4f21+a6a2+4701+87df+19bba58d2018>

Download: [.csv](#) [.tsv](#) [.json](#)

Feedback: [Dieses Mapping in Cocoda bewerten](#) oder [Feedback per E-Mail senden](#) (coli-conc@gbv.de)

Web-Ansicht (JSKOS Viewer)

Kontext-Info

[In Cocoda öffnen](#)

Quelle (RVK)

QR 540 - QR 544 Handelsgewerbe

┆ Q Wirtschaftswissenschaften

┆ QR Gewerbepolitik. Einzelne Wirtschaftszweige

┆ QR 500 - QR 870 Industrie, Bergbau, Handel, Dienstleistungen, Handwerk

↳ QR 540 Allgemeines

↳ QR 541 Geschichte

↳ QR 544 Nahrungsmittel Einzelhandel

[Info](#) [Bezeichnungen](#) [Editorial](#)

Ziel (BK)

83.71 Handel

┆ 7-8 Sozialwissenschaften

┆ 83 Volkswirtschaft

┆ 83.00 Volkswirtschaft: Allgemeines

┆ 83.60 Wirtschaftssektoren: Allgemeines

┆ 83.68 Dienstleistungen: Allgemeines

[Info](#) [Bezeichnungen](#)

<http://uri.gbv.de/terminology/bk/83.71> (URI)

In Arbeit: Subjects API

- **Abfrage der Sacherschließung** zu einem Titel
- **Änderung der Sacherschließung** per API und Webanwendung (ohne WinIBW!)

Vorteile

- Keine PICA/MARC-Kenntnisse notwendig, nur inhaltliches Fachwissen
- Nicht nur für K10plus, auch Repositories, Forschungsdaten, andere Verbünde...

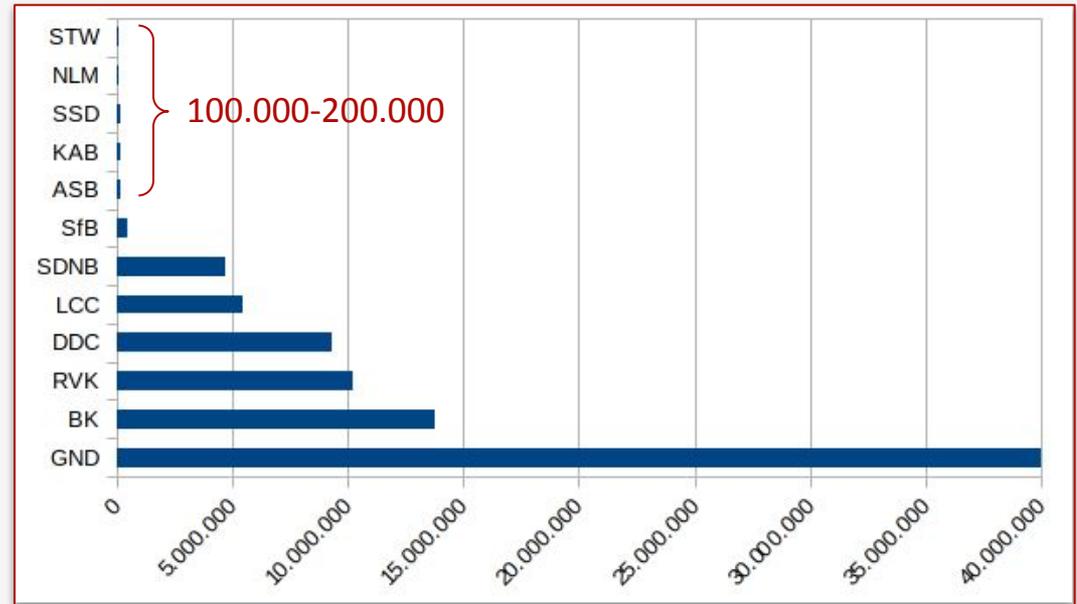
Analyse und Korrekturen vorhandener Erschließung

Bereitstellung von Datenabzügen

- **Sacherschließungsfelder** aus dem K10plus
- Regelmäßige Publikation als Open Data unter <https://doi.org/10.5281/zenodo.7016625>
 - Tabellarisches Format (TSV-Datei)
 - **RDF-Format** (NTripel)
- **Datenbereinigung**
 - Nur syntaktisch korrekte Identifier (Beispiel BK: [0-9][0-9]\.[0-9][0-9])
 - Geplant: Nur tatsächlich vorhandene Normdatensätze/Klassen
 - Konvertierung von Identifier in URIs (83.71 Handel => <http://uri.gbv.de/terminology/bk/83.71>)
- Datenbasis für per Occurrences API und Subjects API

Unterstützte Vokabulare

- Voraussetzungen wie beschrieben (BARTOC & URIs): bislang zwölf Vokabulare
- 32% der K10plus-Titel haben eine dieser Erschließungen
- \emptyset 3,5 Einträge pro Titel



Statistiken und Qualitätsmanagement

- **QA Catalogue** im Testbetrieb für den gesamten K10plus
- Ausschnitte für einzelne Bibliotheken möglich
- Fehlerlisten (PPN) zur Erstellung von Korrekturlisten

Vollständigkeit

Alle Felder und Unterfelder mit der jeweiligen Anzahl und dem Anteil ihres Vorkommens

Bibliothek:

Dokumenttyp: [all](#) · [Druckschriften \(einschließlich Bildbänden\)](#) · [Elektronische Ressource im Fernzugriff](#) · [Tonträger, Videodatenträger, Bildliche Darstellungen](#) · [Elektronische Ressource auf Datenträger](#) · [Mikroform](#) · [Medienkombination](#) · [Handschriftliches Material](#) · [Objekt](#) · [Blindenschriftträger und andere taktile Materialien](#)

Datensätze: **71,757,790**

Feldgruppen

Felder	Anzahl	%
0... PICA+ bibliographic description	71,757,790	100.00
unknown unknown origin	3,840,699	5.35

Felder

	Datensätze		Vorkommen			
	Anzahl	%	Anzahl	von bis	∅	σ
PICA+ bibliographic description						
001A=0200 — Kennung und Datum der Ersterfassung	71,757,790	100.00	71,757,790	1		
001B=0210 — Kennung und Datum der letzten Änderung	71,757,790	100.00	71,758,521	1	2	1,000,00
001Q=000Q — VAL-Warnungen	439	0.00	444	1	3	1,010,01

Anreicherung

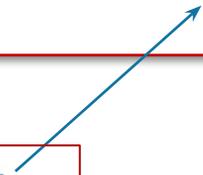
Ergänzung vorhandener Erschließung

coli-ana Anreicherung im Titel

PPN:	766834530 
Titel:	Revolutionary Teamsters : The Minneapolis Truckers' Strikes of 1934
Verfasser:	Palmer, Bryan D.
Ort/Jahr:	Leiden : BRILL, 2013
Sprache/n:	Englisch
Umfang:	Online-Ressource (346 p)
Schriftenreihe:	Historical Materialism Book Series
Anmerkung:	Description based upon print version of record
Mehr zum Titel:	Contents; Acknowledgements; Illustrations; Strike; Chapter Three Uneven and Combine Chapter Five January Thaw; February Cold Auxiliary; Chapter Eight Rebel-Outpost: 190 Chapter Nine The Tribune Alley Plot and the Twelve Toward the July Days; Chapter Thirt Sixteen Martial Law and the Red-Scare; Cha and Solidarity; Chapter Nineteen Mediation's Chapter Twenty-One After 1934: the Reveng the United States, 1928-33; References; Inc
Bibliogr. Zusammenhang:	Print version: Revolutionary Teamsters : The
ISBN:	978-90-04-25420-6 978-90-04-25486-2
Schlagwörter:	*Electronic books
Klassifikation:	Library of Congress Classification: HD5325 .M795 2013 Dewey Decimal Classification: [23] 331.892813883240977657909043 https://www.gbv.de/dms/ddc/766834530-331.892813883240977657909043-v4-k.pdf
Inhalt:	Can workers win? Bryan D. Palmer presents a detailed account of the Minneapolis teamsters' strikes of 1934 to su bad the circumstances
Links zum Titel:	https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=1367809 [Volltext]

PPN: 766834530 - DDC: 331.892813883240977657909043

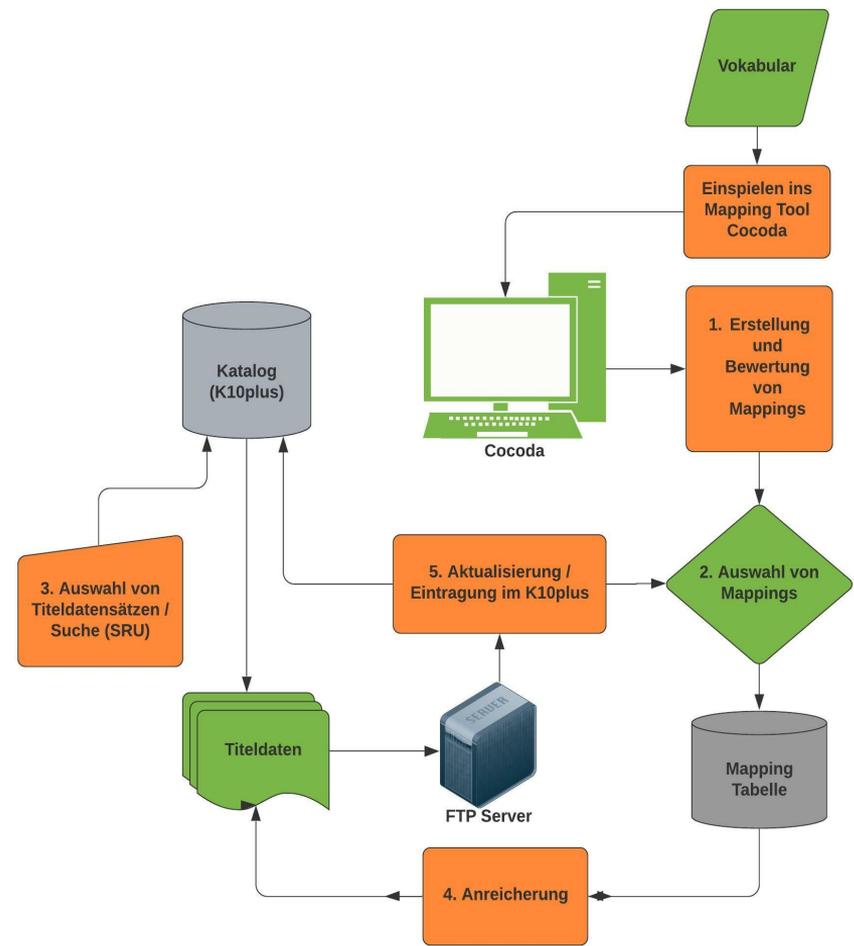
331.89281	Streiks in anderen Industrien als den, anderen Berufe als in den Grundstoffindustrien, dem verarbeitenden Gewerbe, der Bauwirtschaft
388.324	*Lastkraftwagenverkehr
T1--093-099	Einzelne Kontinente, Länder, Ortschaften; außerirdische Welten
T2--776579	Minneapolis
T1--093-099:0901-0905	Zeitabschnitte
T1--09043	*1930-1939



coli-rich Anreicherung

- Erstellung von Mappings in Cocoda
- Überprüfung der Vollständigkeit
- Abzug von Mappings und Titeldaten
- Aktualisierung des K10plus auf Grundlage der Mappings

Über 300.000 Titeldaten wurden mit einer BK-Notation angereichert auf Grundlage von RVK-BK und NSK-BK Mappings.
Nächste Anreicherung: 1 Mio. Titeldaten



coli-rich-Anreicherung im K10plus

K10plusPPN:	1756577099 
Titel:	Diagnostic Pathology: Forensic Autopsy / Mary Ann Sens, Rhome Hughes
Autorin/Autor:	Sens, Mary Ann [Verfasserin/Verfasser]
Beteiligt:	Hughes, Rhome [Mitwirkende/Mitwirkender]
Erschienen:	Philadelphia : Elsevier, 2021
Umfang:	xxi, 571, xxxv pages : [zahlreiche farbige Illustrationen] ; 28 cm
Sprache(n):	Englisch
Anmerkung:	Zugang zu "enhanced digital version" mit Code.
Bibliogr. Zusammenhang:	Erscheint auch als: (Online-Ausgabe)
ISBN:	978-0-323-75617-4
Sonstige Nummern:	OCOLC: 1277170607 
RVK-Notation:	XL 1600 INFO XL 2101 INFO XL 2103 INFO 
Sachgebiete:	Basisklassifikation: 44.72 (Rechtsmedizin)
Sonstige Schlagwörter:	MeSH: Forensic ; Autopsy
Inhaltliche Zusammenfassung:	Intro -- Title Page -- Dedication -- Introduction -- 1. So you want to be a forensic pathologist -- 2. The forensic pathologist -- 3. The forensic pathologist's role in the legal system -- 4. The forensic pathologist's role in the forensic laboratory -- 5. The forensic pathologist's role in the forensic laboratory -- 6. Lightning and environmental deaths -- 7. Every contact leaves a trace -- 8. The forensic pathologist's role in the forensic laboratory -- 9. The forensic pathologist's role in the forensic laboratory -- 10. The forensic pathologist's role in the forensic laboratory -- 11. The forensic pathologist's role in the forensic laboratory -- 12. The perfect murder ... -- 13. Animals and death -- 14. Life from the dead -- 15. The forensic pathologist's role in the forensic laboratory -- 16. The forensic pathologist's role in the forensic laboratory -- 17. The forensic pathologist's role in the forensic laboratory -- 18. On death and dying -- Acknowledgements -- Glossary --

coli-rich-Anreicherung im PICA-Datensatz

```
001@ f0711,2006
001A f03806:29-04-21
001B f0BSZ:21-07-22ft15:17:10.000
001D f03806:29-04-21
001U f0utf8
001X f00
002@ f0Aau
002C faTextfbtxt
002D faohne Hilfsmittel zu benutzenfbn
002E faBandfbnc
003@ f01756577099
0030 faOCoLCf01277170607
004A f0978-0-323-75617-4
007G fiKXPf01756577099
010@ faeng
010E ferda
011@ fa2021
021A faDiagnostic Pathology: Forensic AutopsyfhMary Ann Sens, Rhome Hughes
028A fdMary AnnfaSensfbVerfasserInf4aut
028C fdRhomefaHughesfbMitwirkenderf4ctb
033A fpPhiladelphiafnElsevier
034D faxxi, 571, xxxv pages
034I fa28 cm
034M fa[zahlreiche farbige Illustrationen]
037A faZugang zu "enhanced digital version" mit Code.
039D fiErscheint auch alsfnOnline-AusgabefCISBNf6978-0-323-75618-1
044C fn650fsfaForensicfxPathology
044C fn650fsfaAutopsy
045@/01 f9106409476f844.72§jRechtsmedizinfAcoli-conc RVK->BKfAhttps://coli-conc.gbv.de/api/mappings/d415aba4-14c2-4a9c-822a-1a589787545d
045R f91271582376f8XL 1600: X-Y§jMedizin / XL§jRechtsmedizin / XL 1400-XL 6099§jAllgemeine Rechtsmedizin / XL 1600-XL 2199§jAllgemeine r
1699§jAllgemeine rechtsmedizinische Diagnostik und Methodik (z.B. Tod und Leichenveränderungen) / AllgemeinesfADE-14/sred
045R f9127708291Xf8XL 2101: X-Y§jMedizin / XL§jRechtsmedizin / XL 1400-XL 6099§jAllgemeine Rechtsmedizin / XL 1600-XL 2199§jAllgemeine r
2199§jTodesfeststellung, Todeszeitbestimmung / XL 2101-XL 2128§jA. Formalschlüssel / Hand- und Lehrbücher, einführende Werke
045R f91277082944f8XL 2103: X-Y§jMedizin / XL§jRechtsmedizin / XL 1400-XL 6099§jAllgemeine Rechtsmedizin / XL 1600-XL 2199§jAllgemeine r
2199§jTodesfeststellung, Todeszeitbestimmung / XL 2101-XL 2128§jA. Formalschlüssel / Kompendien, Repetitorien, Leitfäden, Kurzdarstellun
047A fa***ACHTUNG***Beim Kauf dieser Ausgabe wird ein Online-Zugang für Privatpersonen angeboten. Dieser darf nur von Einzelpersonen mit
```

In Arbeit: coli-rich Webanwendung

**coli-rich-web**← zurück zur coli-conc Webseite ISIL: 12345▼ EN
DE

Semi-automatische Eintragung von Sacherschließungsdaten in PICA-Datenbanken.

Mit folgendem Formular können Titeldaten aus PICA-Datenbanken abgerufen werden, um diese mit Sacherschließungsdaten (auf Basis von coli-conc Mappings) anzureichern.

Datenbankauswahl K10plus-Verbundkatalog ▼

Titel laden

Titelangaben

PPN

Titel: Estimating Ore Grade Using Evolutionary Machine Learning Modelsinfo / Ehteram, Mohammad. - 1st ed. 2023. - Singapore : Springer Nature Singapore, 2023

Sacherschließung DDC: 549
bicsc: RBGG ; bisacsh: SCIO48000
LOC-SH: Mineralogy. ; Machine learning. ; Artificial intelligence--Data processing. ; Artificial intelligence.

Mögliche Anreicherung

Vokabular	Notation	Quelle	<input type="checkbox"/>
BK	38.30 Mineralogie	DDC: 549 Mineralogie = BK: 38:30 Mineralogie (Autor, Datum) Details DDC: 549 Mineralogie -> BK: 38:30 Mineralogie (Autor, Datum) Details	<input checked="" type="checkbox"/>

Eintragung erfolgreich! ✓

coli-rich-web ist Teil des coli-conc Projekts. Der Quellcode und die technische Dokumentation werden über [GitHub](#) bereitgestellt. Dieses Tool ist noch in der Entwicklungsphase.

coli-conc ist ein Projekt der [Verbundzentrale des GBV \(VZG\)](#). Es wird gefördert von der Deutschen Forschungsgemeinschaft (DFG) [2015-2019](#) und [2021-2023](#).

Impressum | Barrierefreiheit | Datenschutz

coli-conc-Mappings im DA-3

Mappings im Konkordanz-Register

The Mapping-Editor interface shows two empty boxes for mapping concepts. Below is a table of mappings:

Konkordanzen		Suche		Navigator	
Mappings					
C Konkordanz-Register					
RVK	VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen	<	BK	44.41 Pharmazeutische Biologie	David-Benja... 2022-11-15
RVK	VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen	~	BK	44.73 Geomedizin	Melanie Jan... 2022-04-22
L Lokal keine Ergebnisse </>					
W Wikidata-Mappings keine Ergebnisse </>					

Anzeige im DA-3 als Vorschläge

The DA-3 interface shows a list of suggestions:

- BK
- 44.41 Pharmazeutische Biologie @ddc-narrow/cc, @rvk-narrow/cc
- 44.73 Geomedizin @rvk-related/cc

Quelle: - via RVK VW 6100 «Allgemeines» (Mappingtyp: verwandt) (Melanie Janßen, 2022)

[\[coli-conc\]](#)

Link zu JSKOS Viewer,
dort Link auf Cocoda

Bearbeitung von Mappings in Cocoda

Neues Mapping erstellen zum Ergänzen

RVK Mapping-Editor BK

VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen < 44.39 Toxikologie

Konkordanz-Register: nicht gespeichert V Chemie und Pharmazie (David-Benjamin Rc)

Übernahme in den DA-3

Konkordanzen Suche Navigator

Mappings

RVK	VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen	BK	44.41 Pharmazeutische Biologie	David-Benja...	2022-11-15	+1
RVK	VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen	< BK	44.39 Toxikologie	David-Benja...	2023-05-17	±
RVK	VW 6100 - VW 6109 Nutzpflanzen der Tropen und Subtropen	~ BK	44.73 Geomedizin	Melanie Jan...	2022-04-22	±

Vorschläge Status | Rohdaten | Einstellungen #

Filtern Aktualisieren Erweitern

BK

44.39 Toxikologie @rvk-narrow/cc

Quelle: via RVK VW 6100 «Allgemeines» (Mappingtyp: eng) (David-Benjamin Rohrer, 2023) [coli-conc]

44.41 Pharmazeutische Biologie @ddc-narrow/cc, @rvk-narrow/cc

44.73 Geomedizin @rvk-related/cc

coli-conc Mappings in der WinIBW

The screenshot shows the WinIBW 3.7 interface with a list of records on the left and a 'Passende Mappings' dialog box on the right. The records list includes:

- 001@ f02008, 2018
- 001A f0KALB:17-06-19
- 001B f0KALB:17-06-19ft:16:04:41.000
- 001D f0KALB:17-06-19
- 001V f0utf8
- 001X f00
- 002@ f0Aau
- 002C faTextfbtxt
- 002D faohne Hilfsmittel zu benutzenfbn
- 002E faBandfbnc
- 003@ f01667549030
- 004A f0978-3-608-96167-6
- 007G fikXPf01667549030
- 010@ fager
- 010E ferda
- 011@ fa2019
- 019@ faXA-DE-BW
- 021A faAmerika verstehenfdGeschichte, Politik und Kultur der US
- 028A f9077526112f8Gerste, Ronald D. *1957-* ; ID: gnd/115872566
- 032@ faDritte, durchgesehene Auflage
- 033A fpStuttgartf8Klett-Cotta
- 034D fa202 Seiten
- 034M faKarten
- 041A/00 f9106076612f8USA ; ID: gnd/4078704-7
- 041A/01 f9104152206f8Politik ; ID: gnd/4046514-7
- 041A/02 f9104470348f8Kultur ; ID: gnd/4125698-0
- 041A/03 fzGeschichte
- 041A/09 faDE-101
- 044Z/00 fihisfaUSA
- 044Z/00 fikfGeschichte

The 'Passende Mappings' dialog box displays the following mappings:

- GND 4125698-0 --> LCSH sh85026423 [DNB,2016]
- GND 4046514-7 --> LCSH sh85104466 [DNB,2016]
- GND 4125698-0 --> LCSH sh85034755 [DNB,2016]
- RVK AA 27000 --> GND 4078704-7 [UB Regensburg,...]

Buttons: OK, Abbrechen

Anreicherung paralleler Ausgaben

- Vokabulare: GND, STW, BK, RVK, DDC, freie SW und sonstige Notationen
- Aktuelle Anzahl der angereicherten Titelsätze: **469.385 Titel** (Stand: 28.04.2023)
- In Onlineausgaben (Monografien) mit Verlinkung zu anderen Ausgaben (Routine)
- Neu: In anderen Medienarten, im Test wurden ca. 130.000 Titel angereichert

Lauf vom 08.03.2023	
Ergänzung mit	Anzahl Titel
DDC (5010)	4.021
RVK (5090)	7.290
BK (5301)	13.212
Notationen (5060)	364
STW (5200ff.)	3.237
freie SW (5520)	1.574
GND (5100ff., 5550ff., 5580ff.)	18.742
Summe Titel brutto	48.440
Summe Titel Netto	23.944

RVK-Anreicherung aus Lokaldaten (Exemplare)

- Württemberg. Landesbibliothek:
Anzahl der Angereicherten Titelsätze: 17.238 Titel (Stand: 27.04.2023)
- Staats- und Universitätsbibliothek Hamburg:
analog zur WLB, derzeit in Prüfung

Ausblick

weitere Möglichkeiten und Pläne

Geplant: Weitere Anreicherungen

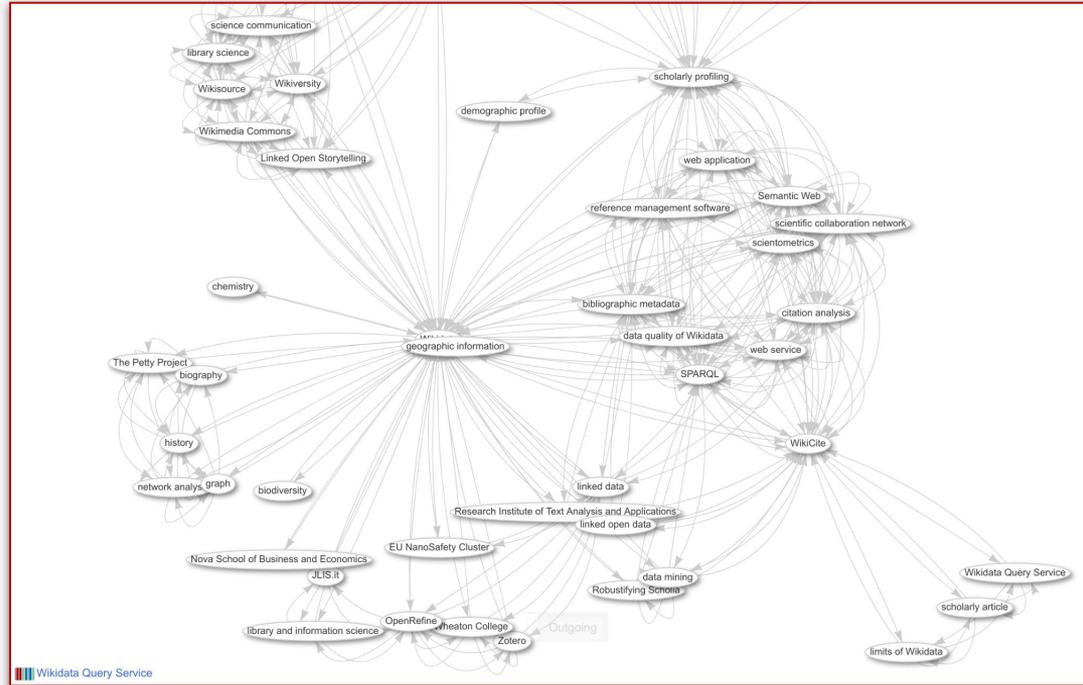
- Optimierung bestehende Verfahren (parallele Ausgaben und Exemplare)
- Auswertung der ISBN-Nummern paralleler Ausgaben auch ohne Verlinkung
- Auswertung der Culturegraph-Werkbündel
- Regelmäßige Hebungen von lokalen Sacherschließungselementen für interessierte Bibliotheken

Aktuell: Automatische Erschließung

- Heuristische Algorithmen zur Ermittlung wahrscheinlicher Sacherschließung
 - Basierend auf vorhandenen Daten (Titel, Autor:innen, Volltexte...)
 - Tendenziell (aber nicht grundsätzlich) schlechter als intellektuelle Erschließung
 - Momentan spannende Entwicklung (LLM, KI...)
-
- Umsetzung durch Bibliotheken je nach Bedarf und Ressourcen (z.B. DNB & ZBW)
 - Eintragung der Ergebnisse im K10plus
 - PICA-Unterfelder \$k und \$v: Verfahren, Konfidenzwert und Datum der Generierung
 - Weitergabe in MARC 883 Metadata Provenance

Forschung: K10plus Knowledge Graph

- Triple-Store aller Informationen in RDF
- Ermöglicht Analysen wie z.B. “Welche Bibliotheken haben Titel aus den 1920er Jahren zu einem Fachgebiet?”



Zusammenfassung

Die VZG bietet datenorientierte Sacherschließung

- Maschinenlesbare Vokabulare (BARTOC, URIs, JSKOS...)
- Web-Schnittstellen (Vokabulare, Mappings, [Co-]Occurrences, Subjects...)
- Web-Anwendungen (Vokabular-Browser, Cocoda, DA-3, coli-ana, QA Catalogue...)
- Datenbankabzüge bereinigter Sacherschließung (TSV, JSON & RDF)
- Statistiken & Anreicherungen der K10plus-Datenbank
auf Grundlage vorhandener Sacherschließung
- Infrastruktur für (aber keine eigene) automatische Erschließung
- Zusammenführung der Daten auch Knowledge Graph in RDF

Fragen?

Warum macht die VZG keine automatische Erschließung?

- Begrenzte Ressourcen
 - Erschließung ist (neben Sammeln und zugänglich Machen) eine der wesentlichen Aufgabe von Bibliotheken
 - Kontrollierte vs. probabilistische Anreicherung
 - Heterogene Metadaten erfordern Bestandsabhängige Verfahren
-
- Ausnahme: Vorschlagsfunktion für die intellektuelle Erschließung
 - Ausnahme: KI-Verfahren im Rahmen des Retrieval (in Entwicklung)